The Synthesis of Two-Terminal Switching Circuits

By CLAUDE. E. SHANNON

PART I: GENERAL THEORY

1. Introduction

THE theory of switching circuits may be divided into two major divisions, analysis and synthesis. The problem of analysis, determining the manner of operation of a given switching circuit, is comparatively simple. The inverse problem of finding a circuit satisfying certain given operating conditions, and in particular the *best* circuit is, in general, more difficult and more important from the practical standpoint. A basic part of the general synthesis problem is the design of a two-terminal network with given operating characteristics, and we shall consider some aspects of this problem.

Switching circuits can be studied by means of Boolean Algebra.[1,2] This is a branch of mathematics that was first investigated by George Boole in connection with the study of logic, and has since been applied in various other fields, such as an axiomatic formulation of Biology,[3] the study of neural networks in the nervous system,[4] the analysis of insurance policies,[5] probability and set theory, etc.

Perhaps the simplest interpretation of Boolean Algebra and the one closest to the application to switching circuits is in terms of propositions. A letter X, say, in the algebra corresponds to a logical proposition. The sum of two letters $X + Y$ represents the proposition "X or Y" and the product XY represents the proposition "X and Y". The symbol X' is used to represent the negation of proposition X, i.e. the proposition "not X". The constants 1 and 0 represent truth and falsity respectively. Thus $X + Y = 1$ means X or Y is true, while $X + YZ' = 0$ means X or (Y and the contradiction of Z) is false.

The interpretation of Boolean Algebra in terms of switching circuits[6,8,9,10] is very similar. The symbol X in the algebra is interpreted to mean a make (front) contact on a relay or switch. The negation of X, written X', represents a break (back) contact on the relay or switch. The constants 0 and 1 represent closed and open circuits respectively and the combining operations of addition and multiplication correspond to series and parallel connections of the switching elements involved. These conventions are shown in Fig. 1. With this identification it is possible to write an algebraic

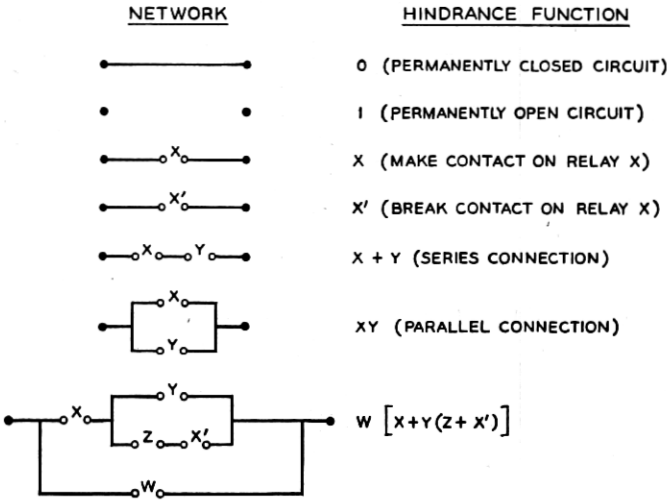


Fig. 1—Hindrance functions for simple circuits.

expression corresponding to a two-terminal network. This expression will involve the various relays whose contacts appear in the network and will be called the hindrance or hindrance function of the network. The last network in Fig. 1 is a simple example.

Boolean expressions can be manipulated in a manner very similar to ordinary algebraic expressions. Terms can be rearranged, multiplied out, factored and combined according to all the standard rules of numerical algebra. We have, for example, in Boolean Algebra the following identities:

$$0 + X = X$$

$$0 \cdot X = 0$$

$$1 \cdot X = X$$

$$X + Y = Y + X$$

$$XY = YX$$

$$X + (Y + Z) = (X + Y) + Z$$

$$X(YZ) = (XY)Z$$

$$X(Y + Z) = XY + XZ$$

The interpretation of some of these in terms of switching circuits is shown in Fig. 2.

There are a number of further rules in Boolean Algebra which allow

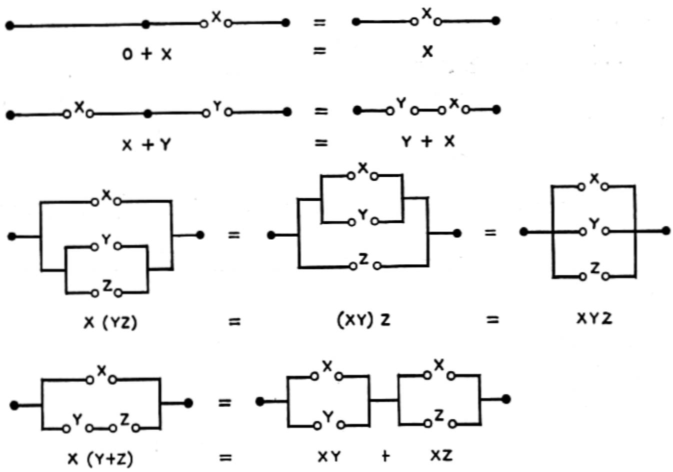


Fig. 2—Interpretation of some algebraic identities.

simplifications of expressions that are not possible in ordinary algebra. The more important of these are:

$$X = X + X = X + X + X = \text{etc.}$$

$$X = X \cdot X = X \cdot X \cdot X = \text{etc.}$$

$$X + 1 = 1$$

$$X + YZ = (X + Y)(X + Z)$$

$$X + X' = 1$$

$$X \cdot X' = 0$$

$$(X + Y)' = X'Y'$$

$$(XY)' = X' + Y'$$

The circuit interpretation of some of these is shown in Fig. 3. These rules make the manipulation of Boolean expressions considerably simpler than ordinary algebra. There is no need, for example, for numerical coefficients or for exponents, since $nX = X^n = X$.

By means of Boolean Algebra it is possible to find many circuits equivalent in operating characteristics to a given circuit. The hindrance of the given circuit is written down and manipulated according to the rules. Each different resulting expression represents a new circuit equivalent to the given one. In particular, expressions may be manipulated to eliminate elements which are unnecessary, resulting in simple circuits.

Any expression involving a number of variables $X_1, X_2, \cdots, X_n$ is called a *function* of these variables and written in ordinary function notation, $f(X_1, X_2, \cdots, X_n)$. Thus we might have $f(X, Y, Z) = X + Y'Z + XZ'$. In Boolean Algebra there are a number of important general theorems which hold for any function. It is possible to *expand* a function about one or more of its arguments as follows:

$$f(X_1, X_2, \cdots, X_n) = X_1 f(1, X_2, \cdots, X_n) + X' f(0, X_2, \cdots, X_n)$$

This is an expansion about X_1. The term $f(1, X_2, \cdots, X_n)$ is the function

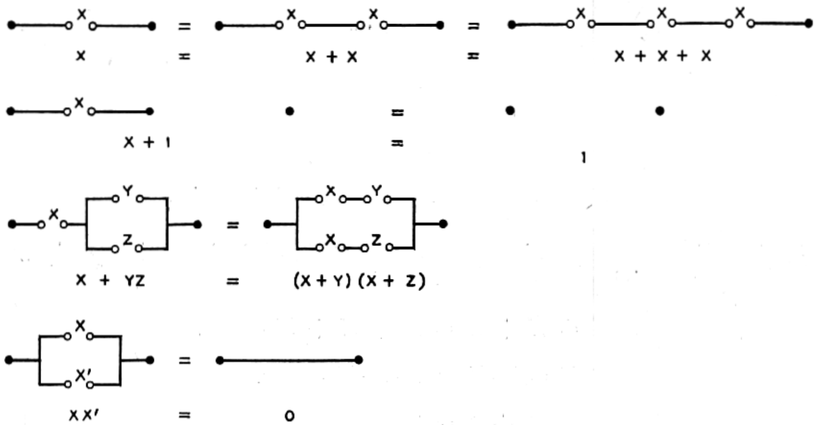


Fig. 3—Interpretation of some special Boolean identities.

$f(X_1, X_2, \cdots, X_n)$ with 1 substituted for X, and 0 for X', and conversely for the term $f(0, X_2, \cdots, X_n)$. An expansion about X_1 and X_2 is:

$$f(X_1, X_2, \cdots, X_n) = X_1 X_2 f(1, 1, X_3, \cdots, X_n) + X_1 X_2' f(1, 0, X_3, \cdots, X_n) + X_1' X_2 f(0, 1, X_3, \cdots, X_n) + X_1' X_2' f(1, 1, X_3, \cdots, X_n)$$

This may be continued to give expansions about any number of variables. When carried out for all n variables, f is written as a sum of 2^n products each with a coefficient which does not depend on any of the variables. Each coefficient is therefore a constant, either 0 or 1.

There is a similar expansion whereby f is expanded as a product:

$$\begin{aligned} f(X_1, X_2, \cdots, X_2) & \\ &= [X_1 + f(0, X_2, \cdots, X_n)]\,[X_1' + f(1, X_2, \cdots, X_n)] \\ &= [X_1 + X_2 + f(0, 0, \cdots, X_n)]\,[X_1 + X_2' + f(0, 1, \cdots, X_n)] \\ &\quad [X_1' + X_2 + f(1, 0, \cdots, X_n)]\,[X_1' + X_2' + f(1, 1, \cdots, X_n)] \\ &= \text{etc.} \end{aligned}$$

The following are some further identities for general functions:

$$X + f(X, Y, Z, \cdots) = X + f(0, Y, Z, \cdots)$$

$$X' + f(X, Y, Z, \cdots) = X' + f(1, Y, Z, \cdots)$$

$$Xf(X, Y, Z, \cdots) = Xf(1, Y, Z, \cdots)$$

$$X'f(X, Y, Z, \cdots) = X'f(0, Y, Z, \cdots)$$

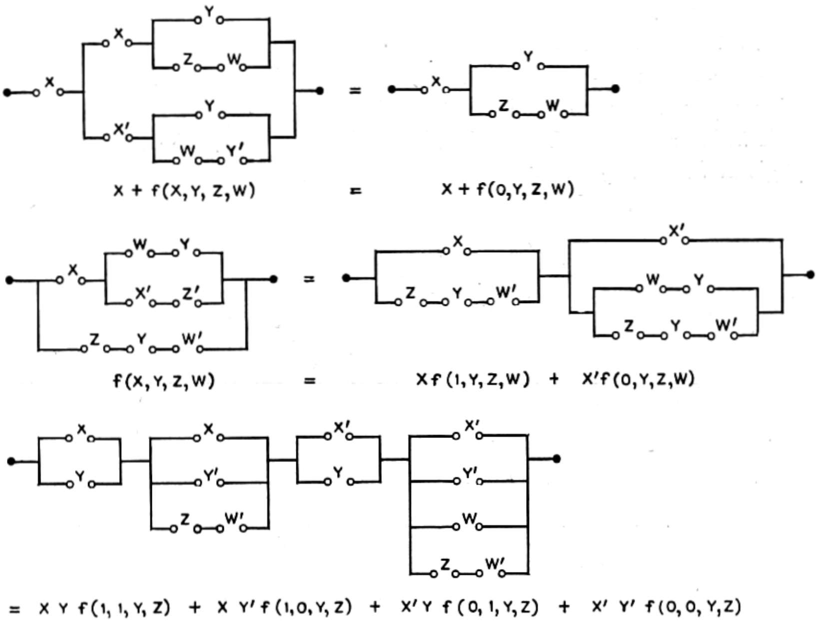


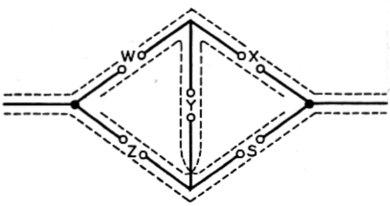
Fig. 4—Examples of some functional identities.

The network interpretations of some of these identities are shown in Fig. 4. A little thought will show that they are true, in general, for switching circuits.

The hindrance function associated with a two-terminal network describes the network completely from the external point of view. We can determine from it whether the circuit will be open or closed for any particular position of the relays. This is done by giving the variables corresponding to operated relays the value 0 (since the make contacts of these are then closed and the break contacts open) and unoperated relays the value 1. For example, with the function $f = W[X + Y(Z + X')]$ suppose relays X and Y operated and Z and W not operated. Then $f = 1[0 + 0(1 + 1)] = 0$ and in this condition the circuit is closed.

A hindrance function corresponds explicitly to a series-parallel type of circuit, i.e. a circuit containing only series and parallel connections. This is because the expression is made up of sum and product operations. There is however, a hindrance function representing the operating characteristics (conditions for open or closed circuits between the two terminals) for any network, series-parallel or not. The hindrance for non-series-parallel networks can be found by several methods of which one is indicated in Fig. 5 for a simple bridge circuit. The hindrance is written as the product of a set of factors. Each factor is the series hindrance of a possible path between the two terminals. Further details concerning the Boolean method for switching circuits may be found in the references cited above.

This paper is concerned with the problem of synthesizing a two-terminal circuit which represents a given hindrance function $f(X_1, \cdots, X_n)$. Since any given function f can be realized in an unlimited number of different



f = (W+X)(Z+S)(W+Y+S)(Z+Y+X)

Fig. 5—Hindrance of a bridge circuit.

ways, the particular design chosen must depend upon other considerations. The most common of these determining criteria is that of economy of elements, which may be of several types, for example:

(1) We may wish to realize our function with the least total number of switching elements, regardless of which variables they represent.

(2) We may wish to find the circuit using the least total number of relay springs. This requirement sometimes leads to a solution different from (1), since contiguous make and break elements may be combined into transfer elements so that circuits which tend to group make and break contacts on the same relay into pairs will be advantageous for (2) but not necessarily for (1).

(3) We may wish to distribute the spring loading on all the relays or on some subset of the relays as evenly as possible. Thus, we might try to find the circuit in which the most heavily loaded relay was as lightly loaded as possible. More generally, we might desire a circuit in which the loading on the relays is of some specified sort, or as near as possible to this given distribution. For example, if the relay X_1

must operate very quickly, while X_2 and X_3 have no essential time limitations but are ordinary U-type relays, and X_4 is a multicontact relay on which many contacts are available, we would probably try to design a circuit for $f(X_1, X_2, X_3, X_4)$ in such a way as, first of all, to minimize the loading on X_1, next to equalize the loading on X_2 and X_3 keeping it at the same time as low as possible, and finally not to load X_4 any more than necessary. Problems of this sort may be called *problems in spring-load distribution.*

Although all equivalent circuits representing a given function f which contain only series and parallel connections can be found with the aid of Boolean Algebra, the most economical circuit in any of the above senses will often not be of this type. The problem of synthesizing non-series-parallel circuits is exceedingly difficult. It is even more difficult to show that a circuit found in some way is the *most* economical one to realize a given function. The difficulty springs from the large number of essentially different networks available and more particularly from the lack of a simple mathematical idiom for representing these circuits.

We will describe a new design method whereby any function $f(X_1, X_2, \cdots, X_n)$ may be realized, and frequently with a considerable saving of elements over other methods, particularly when the number of variables n is large. The circuits obtained by this method will not, in general, be of the series-parallel type, and, in fact, they will usually not even be planar. This method is of interest theoretically as well as for practical design purposes, for it allows us to set new upper limits for certain numerical functions associated with relay circuits. Let us make the following definitions:

$\lambda(n)$ is defined as the least number such that any function of n variables can be realized with not more than $\lambda(n)$ elements.* Thus, any function of n variables can be realized with $\lambda(n)$ elements and at least one function with no less.

$\mu(n)$ is defined as the least number such that given any function f of n variables, there is a two-terminal network having the hindrance f and using not more than $\mu(n)$ elements on the most heavily loaded relay.

The first part of this paper deals with the general design method and the behaviour of $\lambda(n)$. The second part is concerned with the possibility of various types of spring load distribution, and in the third part we will study certain classes of functions that are especially easy to synthesize, and give some miscellaneous theorems on switching networks and functions.

2. Fundamental Design Theorem

The method of design referred to above is based on a simple theorem dealing with the interconnection of two switching networks. We shall first

* An *element* means a make or break contact on one relay. A *transfer element* means a make-and-break with a common spring, and contains two *elements*.

state and prove this theorem. Suppose that M and N (Fig. 6) are two $(n+1)$ terminal networks, M having the hindrance functions U_k $(k = 1, 2, \cdots n)$ between terminals a and k, and N having the functions V_k between b and k. Further, let M be such that $U_{jk} = 1 (j, k = 1, 2, \cdots, n)$. We will say, in this case, that M is a *disjunctive* network. Under these conditions we shall prove the following:

Theorem 1: If the corresponding terminals 1, 2, $\cdots$, n *of* M *and* N *are connected together, then*

$$U_{ab} = \prod_{k=1}^{n} (U_k + V_k) \tag{1}$$

where U_{ab} *is the hindrance from terminal* a *to terminal* b.

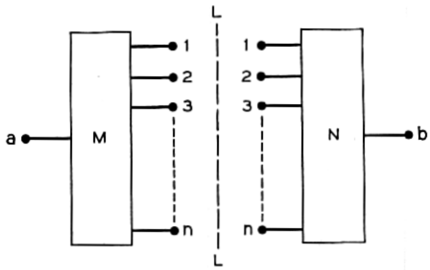


Fig. 6—Network for general design theorem.

Proof: It is known that the hindrance U_{ab} may be found by taking the product of the hindrances of all possible paths from a to b along the elements of the network.[6] We may divide these paths into those which cross the line L once, those which cross it three times, those which cross it five times, etc. Let the product of the hindrances in the first class be W_1, in the second class W_3, etc. Thus

$$U_{ab} = W_1 \cdot W_3 \cdot W_5 \cdots \tag{2}$$

Now clearly

$$W_1 = \prod_{1}^{n} (U_k + V_k)$$

and also

$$W_3 = W_5 = \cdots = 1$$

since each term in any of these must contain a summand of the type U_{jk} which we have assumed to be 1. Substituting in (2) we have the desired result.

The method of using this theorem to synthesize networks may be roughly described as follows: The function to be realized is written in the form of a product of the type (1) in such a way that the functions U_k are the same for a large class of functions, the V_k determining the particular one under consideration. A basic disjunctive network M is constructed having the functions U_k between terminals a and k. A network N for obtaining the functions V_k is then found by inspection or according to certain general rules. We will now consider just how this can be done in various cases.

3. Design of Networks for General Functions—Behavior of $\lambda(n)$.

a. Functions of One, Two and Three Variables:

Functions of one or two variables may be dismissed easily since the number of such functions is so small. Thus, with one variable X, the possible functions are only:

$$0, 1, X, X'$$

and obviously $\lambda(1) = 1$, $\mu(1) = 1$.

With two variables X and Y there are 16 possible functions:

$$0 \quad X \quad Y \qquad XY \qquad XY' \qquad X'Y \qquad X'Y' \qquad XY' + X'Y$$

$$1 \quad X' \quad Y' \qquad X + Y \qquad X + Y' \qquad X' + Y \qquad X' + Y' \qquad XY + X'Y'$$

so that $\lambda(2) = 4$, $\mu(2) = 2$.

We will next show that any function of three variables $f(X, Y, Z)$ can be realized with not more than eight elements and with not more than four from any one relay. Any function of three variables can be expanded in a product as follows:

$$f(X, Y, Z) = [X + Y + f(0, 0, Z)][X + Y' + f(0, 1, Z)]$$
$$[X' + Y + f(1, 0, Z)]\,[X' + Y' + f(1, 1, Z)].$$

In the terminology of Theorem 1 we let

$$U_1 = X + Y \qquad V_1 = f(0, 0, Z)$$
$$U_2 = X + Y' \qquad V_2 = f(0, 1, Z)$$
$$U_3 = X' + Y \qquad V_3 = f(1, 0, Z)$$
$$U_4 = X' + Y' \qquad V_4 = f(1, 1, Z)$$

so that

$$U_{ab} = f(X, Y, Z) = \prod_{k=1}^{4} (U_k + V_k)$$

The above U_k functions are realized with the network M of Fig. 7 and it is easily seen that $U_{jk} = 1$ $(j, k = 1, 2, 3, 4)$. The problem now is to construct a second network N having the V_k functions V_1, V_2, V_3, V_4. Each of these is a function of the one variable Z and must, therefore, be one of the four possible functions of one variable:

$$0, 1, Z, Z'.$$

Consider the network N of Fig. 8. If any of the V's are equal to 0, connect the corresponding terminals of M to the terminal of N marked 0; if any are equal to Z, connect these terminals of M to the terminal of N marked Z, etc. Those which are 1 are, of course, not connected to anything. It is clear from Theorem 1 that the network thus obtained will realize the function $f(X, Y, Z)$. In many cases some of the elements will be superfluous, e.g., if one of the V_i is equal to 1, the element of M connected to terminal i can

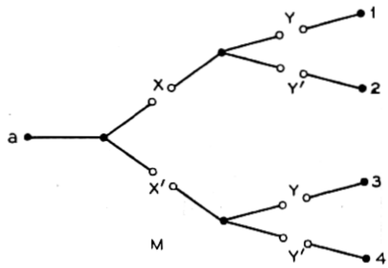


Fig. 7—Disjunctive tree with two bays.

be eliminated. At worst M contains six elements and N contains two. The variable X appears twice, Y four times and Z twice. Of course, it is completely arbitrary which variables we call X, Y, and Z. We have thus proved somewhat more than we stated above, namely,

Theorem 2: Any function of three variables may be realized using not more than 2, 2, *and* 4 *elements from the three variables in any desired order. Thus* $\lambda(3) \leq 8$, $\mu(3) \leq 4$. *Further, since make and break elements appear in adjacent pairs we can obtain the distribution* 1, 1, 2, *in terms of transfer elements.*

The theorem gives only upper limits for $\lambda(3)$ and $\mu(3)$. The question immediately arises as to whether by some other design method these limits could be lowered, i.e., can the $\leq$ signs be replaced by $<$ signs. It can be shown by a study of special cases that $\lambda(3) = 8$, the function

$$X \oplus Y \oplus Z = X(YZ + Y'Z') + X'(YZ' + Y'Z)$$

requiring eight elements in its most economical realization. $\mu(3)$, however, is actually 3.

It seems probable that, in general, the function

$$X_1 \oplus X_2 \oplus \cdots \oplus X_n$$

requires $4(n - 1)$ elements, but no proof has been found. Proving that a certain function cannot be realized with a small number of elements is somewhat like proving a number transcendental; we will show later that almost all* functions require a large number of elements, but it is difficult to show that a particular one does.

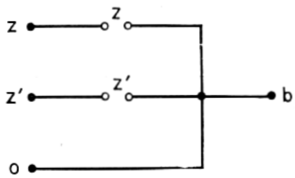


Fig. 8—Network giving all functions of one variable.

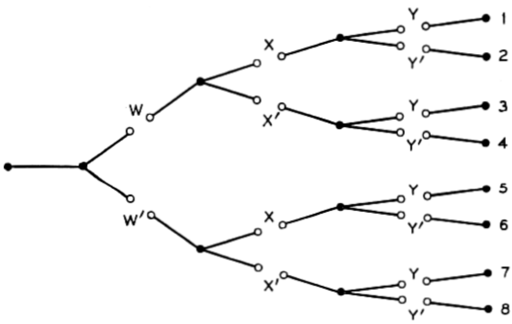


Fig. 9—Disjunctive tree with three bays.

b. Functions of Four Variables:

In synthesizing functions of four variables by the same method, two courses are open. First, we may expand the function as follows:

$$\begin{aligned} f(W, X, Y, Z) = {} & [W + X + Y + V_1(Z)]\cdot[W + X + Y' + V_2(Z)]\cdot \\ & [W + X' + Y + V_3(Z)]\cdot[W + X' + Y' + V_4(Z)]\cdot \\ & [W' + X + Y + V_5(Z)]\cdot[W' + X + Y' + V_5(Z)]\cdot \\ & [W' + X' + Y + V_7(Z)]\cdot[W' + X' + Y' + V_8(Z)]. \end{aligned}$$

By this expansion we would let $U_1 = W + X + Y$, $U_2 = W + X + Y'$, $\cdots$, $U_8 = W' + X' + Y'$ and construct the M network in Fig. 9. N would

* We use the expression "almost all" in the arithmetic sense: e.g., a property is true of almost all functions of n variables if the fraction of all functions of n variables for which it is not true $\to 0$ as $n \to \infty$.

again be as in Fig. 8, and by the same type of reasoning it can be seen that $\lambda(4) \leq 16$.

Using a slightly more complicated method, however, it is possible to reduce this limit. Let the function be expanded in the following way:

$$f(W, X, Y, Z) = [W + X + V_1(Y, Z)] \cdot [W + X' + V_2(Y, Z)]$$
$$[W' + X + V_3(Y, Z)] \cdot [W' + X' + V_4(Y, Z)].$$

We may use a network of the type of Fig. 7 for M. The V functions are now functions of two variables Y and Z and may be any of the 16 functions:

$$A\begin{cases} 0 \\ 1 \end{cases} \quad B\begin{cases} Y \\ Y' \\ Z \\ Z' \end{cases} \quad C\begin{cases} YZ \\ Y'Z \\ YZ' \\ Y'Z' \end{cases} \quad D\begin{cases} Y + Z \\ Y + Z' \\ Y' + Z \\ Y' + Z' \end{cases} \quad E\begin{cases} Y'Z + YZ' \\ YZ + Y'Z' \end{cases}$$

We have divided the functions into five groups, A, B, C, D and E for later reference. We are going to show that any function of four variables can

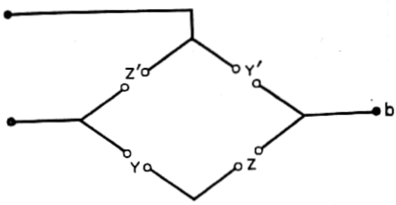


Fig. 10—Simplifying network.

be realized with not more than 14 elements. This means that we must construct a network N using not more than eight elements (since there are six in the M network) for any selection of four functions from those listed above. To prove this, a number of special cases must be considered and dealt with separately:

(1) If all four functions are from the groups, A, B, C, and D, N will certainly not contain more than eight elements, since eight letters at most can appear in the four functions.

(2) We assume now that just one of the functions is from group E; without loss of generality we may take it to be $YZ' + Y'Z$, for it is the other, replacing Y by Y' transforms it into this. If one or more of the remaining functions are from groups A or B the situation is satisfactory, for this function need require no elements. Obviously 0 and 1 require no elements and Y, Y', Z or Z' may be "tapped off" from the circuit for $YZ' + Y'Z$ by writing it as $(Y + Z)(Y' + Z')$. For example, Y' may be obtained with the circuit of Fig. 10. This leaves four elements, certainly a sufficient number for any two functions from A, B, C, or D.

(3) Now, still assuming we have one function, $YZ' + Y'Z$, from E, suppose at least two of the remaining are from D. Using a similar "tapping off" process we can save an element on each of these. For instance, if the functions are $Y + Z$ and $Y' + Z'$ the circuit would be as shown in Fig. 11.

(4) Under the same assumption, then, our worst case is when two of the functions are from C and one from D, or all three from C. This latter case is satisfactory since, then, at least one of the three must be a term of $YZ' + Y'Z$ and can be "tapped off." The former case is bad only when the two functions from C are YZ and $Y'Z'$. It may be seen that the only

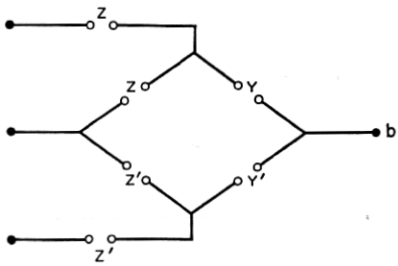


Fig. 11—Simplifying network.

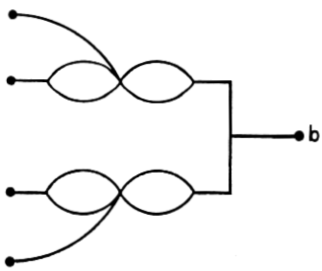


Fig. 12—Simplifying network.

essentially different choices for the function from D are $Y + Z$ and $Y' + Z$. That the four types of functions f resulting may be realized with 14 elements can be shown by writing out typical functions and reducing by Boolean Algebra.

(5) We now consider the cases where two of the functions are from E. Using the circuit of Fig. 12, we can tap off functions or parts of functions from A, B or D, and it will be seen that the only difficult cases are the following: (a) Two functions from C. In this case either the function f is symmetric in Y and Z or else both of the two functions may be obtained from the circuits for the E functions of Fig. 12. The symmetric case is handled in a later section. (b) One is from C, the other from D. There is only one unsymmetric case. We assume the four functions are $Y \oplus Z$, $Y \oplus Z'$, YZ and $Y + Z'$. This gives rise to four types of functions f, which can all be reduced by algebraic methods. This completes the proof.

Theorem 3: Any function of four variables can be realized with not more than 14 *elements.*

c. Functions of More Than Four Variables:

Any function of five variables may be written

$$f(X_1, \cdots, X_5) = [X_5 + f_1(X_1, \cdots, X_4)] \cdot [X_5' + f_2(X_1, \cdots, X_4)]$$

and since, as we have just shown, the two functions of four variables can be realized with 14 elements each, $f(X_1, \cdots X_5)$ can be realized with 30

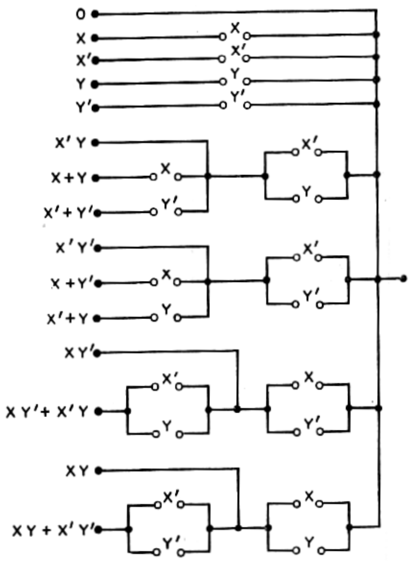


Fig. 13—Network giving all functions of two variables.

Now consider a function $f(X_1, X_2, \cdots, X_n)$ of n variables. For $5 < n \leq 13$ we get the best limit by expanding about all but two variables.

$$f(X_1, X_2, \cdots, X_n) = [X_1 + X_2 + \cdots + X_{n-2} + V_1(X_{n-1}, X_n)] \cdot \cdots \cdot [X_1' + X_2' + \cdots + X_{n-2}' + V_S(X_{n-1}, V_n)] \quad (4)$$

The V's are all functions of the variables X_{n-1}, X_n and may be obtained from the general N network of Fig. 13, in which *every* function of two variables appears. This network contains 20 elements which are grouped into five transfer elements for one variable and five for the other.* The M network for (4), shown in Fig. 14, requires in general $2^{n-1} - 2$ elements. Thus we have:

* Several other networks with the same property as Fig. 13 have been found, but they all require 20 elements.

Theorem 4. $\lambda(n) \leq 2^{n-1} + 18$

d. Upper Limits for $\lambda(n)$ with Large n.

Of course, it is not often necessary to synthesize a function of more than say 10 variables, but it is of considerable theoretical interest to determine as closely as possible the behavior of $\lambda(n)$ for large n.

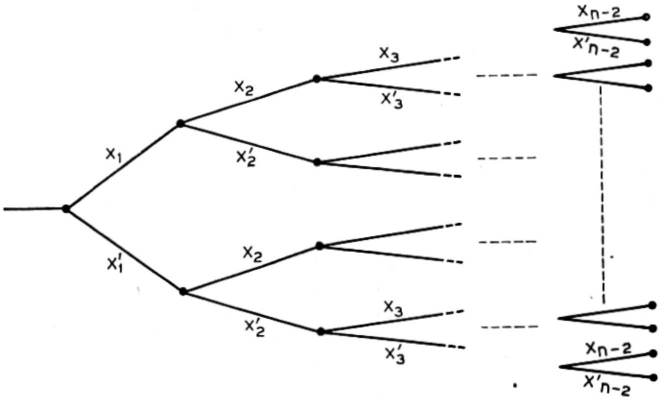


Fig. 14—Disjunctive tree with $(n - 2)$ bays.

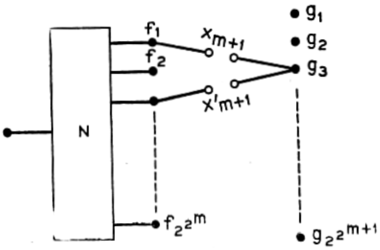


Fig. 15—Network giving all functions of $(m + 1)$ variables constructed from one giving all functions of m variables.

We will first prove a theorem placing limits on the number of elements required in a network analogous to Fig. 13 but generalized for m variables.

Theorem 5. An N network realizing all 2^{2^m} functions of m variables can be constructed using not more than $2 \cdot 2^{2^m}$ elements, i.e., not more than two elements per function. Any network with this property uses at least $(\frac{3}{2} - \epsilon)$ elements per function for any $\epsilon > 0$ with n sufficiently large.

The first part will be proved by induction. We have seen it to be true for $m = 1, 2$. Suppose it is true for some m with the network N of Fig. 15. Any function of $m + 1$ variables can be written

$$g = [X_{m+1} + f_a][X'_{m+1} + f_b]$$

where f_a and f_b involve only m variables. By connecting from g to the corresponding f_a and f_b terminals of the smaller network, as shown typically for g_3, we see from Theorem 1 that all the g functions can be obtained. Among these will be the 2^{2^m} f functions and these can be obtained simply by connecting across to the f functions in question without any additional elements. Thus the entire network uses less than

$$(2^{2^{m+1}} - 2^{2^m})2 + 2 \cdot 2^{2^m}$$

elements, since the N network by assumption uses less than $2 \cdot 2^{2^m}$ and the first term in this expression is the number of added elements.

The second statement of Theorem 7 can be proved as follows. Suppose we have a network, Fig. 16, with the required property. The terminals can be divided into three classes, those that have one or less elements di-

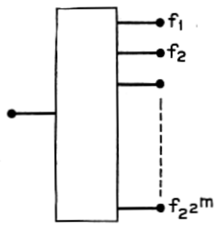


Fig. 16—Network giving all functions of m variables.

rectly connected, those with two, and those with three or more. The first set consists of the functions 0 and 1 and functions of the type

$$(X + f) = X + f_{X=0}$$

where X is some variable or primed variable. The number of such functions is not greater than $2m \cdot 2^{2^{m-1}}$ for there are $2m$ ways of selecting an "X" and then $2^{2^{m-1}}$ different functions $f_{X=0}$ of the remaining $m - 1$ variables. Hence the terminals in this class as a fraction of the total $\to 0$ as $m \to \infty$. Functions of the second class have the form

$$g = (X + f_1)(Y + f_2)$$

In case $X \neq Y'$ this may be written

$$XY + XY'g_{X=1,Y=0} + X'Yg_{X=0,Y=1} + X'Y'g_{X=0,Y=0}$$

and there are not more than $(2m)(2m - 2)[2^{2^{m-2}}]^3$ such functions, again a vanishingly small fraction. In case $X = Y'$ we have the situation shown in Fig. 17 and the XX' connection can never carry ground to another terminal since it is always open as a series combination. The inner ends of these elements can therefore be removed and connected to terminals corresponding to functions of less than m variables according to the equation

$$g = (X + f_1)(X' + f_2) = (X + f_{1X=0})(X' + f_{2X=1})$$

if they are not already so connected. This means that all terminals of the second class are then connected to a vanishingly small fraction of the total terminals. We can then attribute two elements each to these terminals and at least one and one-half each to the terminals of the third group. As these two groups exhaust the terminals except for a fraction which $\rightarrow 0$ as $n \rightarrow \infty$, the theorem follows.

If, in synthesizing a function of n variables, we break off the tree at the $(n - m)$th bay, the tree will contain $2^{n-m+1} - 2$ elements, and we can find an N network with not more than $2^{2^m} \cdot 2$ elements exhibiting every function of the remaining m variables. Hence

$$\lambda(n) \leq 2^{n-m+1} - 2 + 2\ 2^{2^m} < 2^{n-m+1} + 2\ 2^{2^m}$$

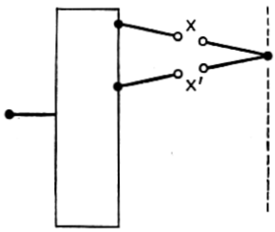


Fig. 17—Possible situation in Fig. 16.

for every integer m. We wish to find the integer $M = M(n)$ minimizing this upper bound.

Considering m as a continuous variable and n fixed, the function

$$f(m) = 2^{n-m+1} + 2^{2^m} \cdot 2$$

clearly has just one minimum. This minimum must therefore lie between m_1 and $m_1 + 1$, where

$$f(m_1) = f(m_1 + 1)$$

i.e.,

$$2^{n-m_1+1} + 2^{2^{m_1}} \cdot 2 = 2^{n-m_1} + 2^{2^{m_1+1}} \cdot 2$$

or

$$2^n = 2^{m_1+1}(2^{2^{m_1+1}} - 2^{2^{m_1}})$$

Now m_1 cannot be an integer since the right-hand side is a power of two and the second term is less than half the first. It follows that to find the *integer* M making $f(M)$ a minimum we must take for M the least integer satisfying

$$2^n \leq 2^{M+1} 2^{2^{M+1}}$$

Thus M satisfies:

$$M + 1 + 2^{M+1} \geq n > M + 2^M \tag{5}$$

This gives:

$n \leq 11$	$M = 2$
$11 < n \leq 20$	$M = 3$
$20 < n \leq 37$	$M = 4$
$37 < n \leq 70$	$M = 5$
$70 < n \leq 135$	$M = 6$
etc.	

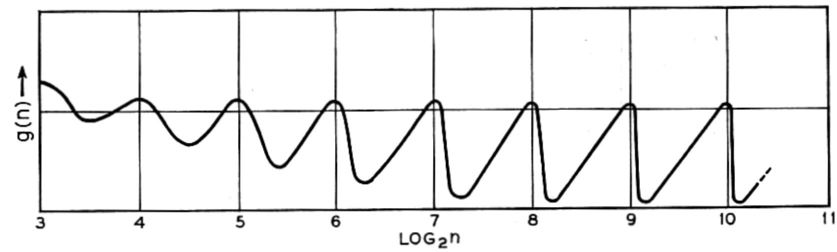


Fig. 18—Behaviour of $g(n)$.

Our upper bound for $\lambda(n)$ behaves something like $\frac{2^{n+1}}{n}$ with a superimposed saw-tooth oscillation as n varies between powers of two, due to the fact that m must be an integer. If we define $g(n)$ by

$$2^{n-M+1} + 2^{2^M}2 = g(n)\,\frac{2^{n+1}}{n},$$

M being determined to minimize the function (i.e., M satisfying (5)), then $g(n)$ varies somewhat as shown in Fig. 18 when plotted against $\log_2 n$. The maxima occur just beyond powers of two, and closer and closer to them as $n \to \infty$. Also, the saw-tooth shape becomes more and more exact. The sudden drops occur just after we change from one value of M to the next. These facts lead to the following:

Theorem 6. (*a*) *For all* n

$$\lambda(n) < \frac{2^{n+3}}{n}.$$

(*b*) *For almost all* n

$$\lambda(n) < \frac{2^{n+2}}{n}.$$

(*c*) *There is an infinite sequence of* n_i *for which*

$$\lambda(n_i) < \frac{2^{n+1}}{n}(1+\epsilon) \qquad \epsilon > 0.$$

These results can be proved rigorously without much difficulty.

e. A Lower Limit for $\lambda(n)$ with Large n.

Up to now most of our work has been toward the determination of upper limits for $\lambda(n)$. We have seen that for all n

$$\lambda(n) < B\frac{2^n}{n}.$$

We now ask whether this function $B\frac{2^n}{n}$ is anywhere near the true value of $\lambda(n)$, or may $\lambda(n)$ be perhaps dominated by a smaller order of infinity, e.g., n^p. It was thought for a time, in fact, that $\lambda(n)$ might be limited by n^2 for all n, arguing from the first few values: 1, 4, 8, 14. We will show that this is far from the truth, for actually $\frac{2^n}{n}$ is the correct order of magnitude of $\lambda(n)$:

$$A\frac{2^n}{n} < \lambda(n) < B\frac{2^n}{n}$$

for all n. A closely associated question to which a partial answer will be given is the following: Suppose we define the "complexity" of a given function f of n variables as the ratio of the number of elements in the most economical realization of f to $\lambda(n)$. Then any function has a complexity lying between 0 and 1. Are most functions simple or complex?

Theorem 7: *For all sufficiently large* n, *all functions of* n *variables excepting a fraction* δ *require at least* $(1-\epsilon)\frac{2^n}{n}$ *elements, where* ϵ *and* δ *are arbitrarily small positive numbers. Hence for large* n

$$\lambda(n) > (1-\epsilon)\frac{2^n}{n}$$

and almost all functions have a complexity $> \frac{1}{4}(1-\epsilon)$. *For a certain sequence* n_i *almost all functions have a complexity* $> \frac{1}{2}(1-\epsilon)$.

The proof of this theorem is rather interesting, for it is a pure existence proof. We do not show that any particular function or set of functions requires $(1-\epsilon)\frac{2^n}{n}$ elements, but rather that it is impossible for all functions to require less. This will be done by showing that there are not enough networks with less than $(1 - \epsilon)\frac{2^n}{n}$ branches to go around, i.e., to represent all the 2^{2^n} functions of n variables, taking account, of course, of the different assignments of the variables to the branches of each network. This is only possible due to the extremely rapid increase of the function 2^{2^n}. We require the following:

Lemma: The number of two-terminal networks with K or less branches is less than $(6K)^K$.

Any two-terminal network with K or less branches can be constructed as follows: First line up the K branches as below with the two terminals a and b.

a.	1—1′
	2—2′
	3—3′
	4—4′
	.
	.
	.
b.	K—K'

We first connect the terminals $a, b, 1, 2, \cdots, K$ together in the desired way. The number of *different* ways we can do this is certainly limited by the number of partitions of $K + 2$ which, in turn, is less than

$$2^{K+1}$$

for this is the number of ways we can put one or more division marks between the symbols $a, 1, \cdots, K, b$. Now, assuming $a, 1, 2, \cdots, K, b$, interconnected in the desired manner, we can connect $1'$ either to one of these terminals or to an additional junction point, i.e., $1'$ has a choice of at most

$$K + 3$$

terminals, $2'$ has a choice of at most $K + 4$, etc. Hence the number of networks is certainly less than

$$2^{K+1}(K + 3)\,(K + 4)\,(K + 5) \cdots (2K + 3)$$

$$< (6K)^K \qquad K \geq 3$$

and the theorem is readily verified for $K = 1, 2$.

We now return to the proof of Theorem 7. The number of functions of n variables that can be realized with $\frac{(1 - \epsilon)2^n}{n}$ elements is certainly less than the number of networks we can construct with this many branches multiplied by the number of assignments of the variables to the branches, i.e., it is less than

$$H = (2n)^{(1-\epsilon)(2^n/n)}\left[6(1-\epsilon)\frac{2^n}{n}\right]^{(1-\epsilon)(2^n/n)}$$

Hence

$$\log_2 H = (1-\epsilon)\frac{2^n}{n}\log 2n + (1-\epsilon)\frac{2^n}{n}\log (1-\epsilon)\frac{2^n}{n}\cdot 6$$

$$= (1-\epsilon)\,2^n + \text{terms dominated by this term for large } n.$$

By choosing n so large that $\frac{\epsilon}{2}2^n$ dominates the other terms of $\log H$ we arrive at the inequality

$$\log_2 H < (1-\epsilon_1)\,2^n$$
$$H < 2^{(1-\epsilon_1)2^n}$$

But there are $S = 2^{2^n}$ functions of n variables and

$$\frac{H}{S} = \frac{2^{(1-\epsilon_1)2^n}}{2^{2^n}} \to 0 \quad \text{as} \quad n \to \infty.$$

Hence almost all functions require more than $(1-\epsilon_1)2^n$ elements.

Now, since for all $n > N$ there is at least one function requiring more than (say) $\frac{1}{2}\frac{2^n}{n}$ elements and since $\lambda(n) > 0$ for $n > 0$, we can say that for *all* n,

$$\lambda(n) > A\frac{2^n}{n}$$

for some constant $A > 0$, for we need only choose A to be the minimum number in the finite set:

$$\frac{1}{2},\quad \frac{\lambda(1)}{\frac{2^1}{1}},\quad \frac{\lambda(2)}{\frac{2^2}{2}},\quad \frac{\lambda(3)}{\frac{2^3}{3}},\ \cdots,\ \frac{\lambda(N)}{\frac{2^N}{N}}$$

Thus $\lambda(n)$ is of the order of magnitude of $\frac{2^n}{n}$. The other parts of Theorem 8 follow easily from what we have already shown.

The writer is of the opinion that almost all functions have a complexity nearly 1, i.e., $> 1-\epsilon$. This could be shown at least for an infinite sequence n_i if the Lemma could be improved to show that the number of networks is less than $(6K)^{K/2}$ for large K. Although several methods have been used in counting the networks with K branches they all give the result $(6K)^K$.

It may be of interest to show that for large K the number of networks is greater than

$$(6K)^{K/4}$$

This may be done by an inversion of the above argument. Let $f(K)$ be the number of networks with K branches. Now, since there are 2^{2^n} functions of n variables and each can be realized with $(1 + \epsilon) \frac{2^{n+2}}{n}$ elements (n sufficiently large),

$$f\left((1 + \epsilon) \frac{2^{n+2}}{n}\right) (2n)^{(1+\epsilon)(2^{n+2}/n)} > 2^{2^n}$$

for n large. But assuming $f(K) < (6K)^{K/4}$ reverses the inequality, as is readily verified. Also, for an infinite sequence of K,

$$f(K) > (6K)^{K/2}$$

Since there is no obvious reason why $f(K)$ should be connected with powers of 2 it seems likely that this is true for all large K.

We may summarize what we have proved concerning the behavior of $\lambda(n)$ for large n as follows. $\lambda(n)$ varies somewhat as $\frac{2^{n+1}}{n}$; if we let

$$\lambda(n) = A_n \frac{2^{n+1}}{n}$$

then, for large n, A_n lies between $\frac{1}{2} - \epsilon$ and $(2 + \epsilon)$, while for an infinite sequence of n, $\frac{1}{2} - \epsilon < A_n < 1 + \epsilon$.

We have proved, incidentally, that the new design method cannot, in a sense, be improved very much. With series-parallel circuits the best known limit* for $\lambda(n)$ is

$$\lambda(n) < 3.2^{n-1} + 2$$

and almost all functions require $(1 - \epsilon) \frac{2^n}{\log_2 n}$ elements.[7] We have lowered the order of infinity, dividing by at least $\frac{n}{\log_2 n}$ and possibly by n. The best that can be done now is to divide by a constant factor ≤ 4, and for some n, ≤ 2. The possibility of a design method which does this seems, however, quite unlikely. Of course, these remarks apply only to a perfectly general design method, i.e., one applicable to *any* function. Many special *classes* of functions can be realized by special methods with a great saving.

* Mr. J. Riordan has pointed out an error in my reasoning in (6) leading to the statement that this limit is actually reached by the function $X_1 \oplus X_2 \oplus \ldots \oplus X_n$, and has shown that this function and its negative can be realized with about n^2 elements. The error occurs in Part IV after equation 19 and lies in the assumption that the factorization given is the best.

PART II: CONTACT LOAD DISTRIBUTION

4. Fundamental Principles

We now consider the question of distributing the spring load on the relays as evenly as possible or, more generally, according to some preassigned scheme. It might be thought that an attempt to do this would usually result in an increase in the total number of elements over the most economical circuit. This is by no means true; we will show that in many cases (in fact, for almost all functions) a great many load distributions may be obtained (including a nearly uniform distribution) while keeping the total number of elements at the same minimum value. Incidentally this result has a bearing on the behavior of $\mu(n)$, for we may combine this result with

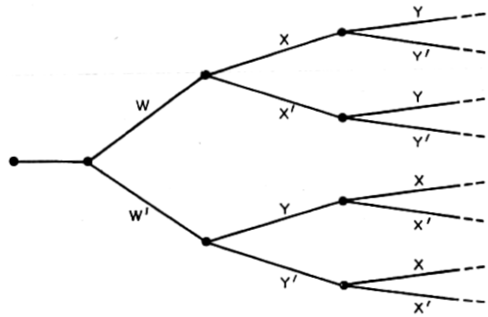


Fig. 19—Disjunctive tree with the contact distribution 1, 3, 3.

preceding theorems to show that $\mu(n)$ is of the order of magnitude of $\frac{2^{n+1}}{n^2}$ as $n \to \infty$ and also to get a good evaluation of $\mu(n)$ for small n.

The problem is rather interesting mathematically, for it involves additive number theory, a subject with few if any previous applications. Let us first consider a few simple cases. Suppose we are realizing a function with the tree of Fig. 9. The three variables appear as follows:

$$W, X, Y \qquad \text{appear}$$
$$2, 4, 8 \qquad \text{times, respectively}\cdot$$

or, in terms of transfer elements*

$$1, 2, 4.$$

Now, W, X, and Y may be interchanged in any way without altering the operation of the tree. Also we can interchange X and Y in the lower branch of the tree only without altering its operation. This would give the distribution (Fig. 19)

$$1, 3, 3$$

* In this section we shall always speak in terms of transfer elements.

A tree with four bays can be constructed with any of the following distributions

W	X	Y	Z	
1,	2,	4,	8	= 1, 2, 4, + 1, 2, 4
1,	2,	5,	7	= 1, 2, 4 + 1, 3, 3
1,	2,	6,	6	= 1, 2, 4 + 1, 4, 2
1,	3,	3,	8	= 1, 2, 4 + 2, 1, 4
1,	3,	4,	7	= 1, 3, 3 + 2, 1, 4
1,	3,	5,	6	= 1, 4, 2 + 2, 1, 4
1,	4,	4,	6	= 1, 3, 3 + 3, 1, 3
1,	4,	5,	5	= 1, 4, 2 + 3, 1, 3

and the variables may be interchanged in any manner. The "sums" on the right show how these distributions are obtained. The first set of numbers represents the upper half of the tree and the second set the lower half. They are all reduced to the sum of sets 1, 2, 4 or 1, 3, 3 in some order, and these sets are obtainable for trees with 3 bays as we already noted. In general it is clear that if we can obtain the distributions

$$a_1, a_2, a_3, \cdots, a_n$$

$$b_1, b_2, b_3, \cdots, b_n$$

for a tree with n bays then we can obtain the distribution

$$1, a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n$$

for a tree with $n + 1$ bays.

Now note that all the distributions shown have the following property: any one may be obtained from the first, 1, 2, 4, 8, by moving one or more units from a larger number to a smaller number, or by a succession of such operations, without moving any units to the number 1. Thus 1, 3, 3, 8 is obtained by moving a unit from 4 to 2. The set 1, 4, 5, 5 is obtained by first moving two units from the 8 to the 2, then one unit to the 4. Furthermore, every set that may be obtained from the set 1, 2, 4, 8 by this process appears as a possible distribution. This operation is somewhat analogous to heat flow—heat can only flow from a hotter body to a cooler one just as units can only be transferred from higher numbers to lower ones in the above.

These considerations suggest that a disjunctive tree with n bays can be constructed with any load distribution obtained by such a flow from the initial distribution

$$1, 2, 4, 8, \cdots, 2^{n-1}$$

We will now show that this is actually the case.

First let us make the following definition: The symbol $(a_1, a_2, \cdots, a_n)$ represents any set of numbers $b_1, b_2, \cdots, b_n$ that may be obtained from the set $a_1, a_2, \cdots, a_n$ by the following operations:

1. Interchange of letters.

2. A flow from a larger number to a smaller one, no flow, however, being allowed to the number 1. Thus we would write

$$1, 2, 4, 8 = (1, 2, 4, 8)$$

$$4, 4, 1, 6 = (1, 2, 4, 8)$$

$$1, 3, 10, 3, 10 = (1, 2, 4, 8, 12)$$

but $2, 2 \neq (1, 3)$. It is possible to put the conditions that

$$b_1, b_2, \cdots, b_n = (a_1, a_2, \cdots, a_n) \tag{6}$$

into a more mathematical form. Let the a_i and the b_i be arranged as non-decreasing sequences. Then a necessary and sufficient condition for the relation (6) is that

(1) $$\sum_{i=1}^{s} b_i \geq \sum_{1}^{s} a_i \qquad s = 1, 2, \cdots, n,$$

(2) $$\sum_{1}^{n} b_i = \sum_{1}^{n} a_i, \quad \text{and}$$

(3) There are the same number of 1's among the a_i as among the b_i. The necessity of (2) and (3) is obvious. (1) follows from the fact that if a_i is non-decreasing, flow can only occur toward the left in the sequence

$$a_1, a_2, a_3, \cdots, a_n$$

and the sum $\sum_{1}^{s} a_i$ can only increase. Also it is easy to see the sufficiency of the condition, for if $b_1, b_2, \cdots, b_n$ satisfies (1), (2), and (3) we can get the b_i by first bringing a_1 up to b_1 by a flow from the a_i as close as possible to a_1 (keeping the "entropy" low by a flow between elements of nearly the same value), then bringing a_2 up to b_2 (if necessary) etc. The details are fairly obvious.

Additive number theory, or the problem of decomposing a number into the sum of numbers satisfying certain conditions, (in our case this definition is generalized to "sets of numbers") enters through the following Lemma:

Lemma: If $a_1, a_2, \cdots, a_n = (2, 4, 8, \cdots, 2^n)$ then we can decompose the a_i into the sum of two sets

$$a_i = b_i + c_i$$

such that

$$b_1, b_2, \cdots, b_n = (1, 2, 4, \cdots, 2^{n-1})$$

and

$$c_1, c_2, \cdots, c_n = (1, 2, 4, \cdots, 2^{n-1})$$

We may assume the a_i arranged in a non-decreasing sequence, $a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n$. In case $a_1 = 2$ the proof is easy. We have

$$\begin{array}{ll} 1, 2, 4, \cdots, 2^{n-1} & \text{B} \\ 1, 2, 4, \cdots, 2^{n-1} & \text{C} \\ \hline 2, 4, 8, \cdots, 2^{n} & \text{A} \end{array}$$

and a flow has occurred in the set

$$4, 8, 16, \cdots, 2^n$$

to give $a_2, a_3, \cdots, a_n$. Now any permissible flow in C corresponds to a permissible flow in either A or B since if

$$c_j = a_j + b_j > c_i = a_i + b_i$$

then either
$$a_j > a_i \quad \text{or} \quad b_j > b_i$$

Thus at each flow in the sum we can make a corresponding flow in one or the other of the summands to keep the addition true.

Now suppose $a_1 > 2$. Since the a_i are non-decreasing

$$(n - 1)\, a_2 \leq (2^{n+1} - 2) - a_1 \leq 2^{n+1} - 2 - 3$$

Hence

$$a_2 - 1 \leq \frac{2^{n+1} - 5}{n - 1} - 1 \leq 2^{n-1}$$

the last inequality being obvious for $n \geq 5$ and readily verified for $n < 5$. This shows that $(a_1 - 1)$ and $(a_2 - 1)$ lie between some powers of two in the set

$$1, 2, 4, \cdots, 2^{n-1}$$

Suppose

$$2^{q-1} < (a_1 - 1) \leq 2^q$$

$$2^{p-1} < (a_2 - 1) \leq a^p \qquad q \leq p \leq (n - 1).$$

Allow a flow between 2^q and 2^{q-1} until one of them reaches $(a_1 - 1)$, the other (say) R; similarly for $(a_2 - 1)$ the other reaching S. As the start toward our decomposition, then, we have the sets (after interchanges)

$$\begin{array}{ll|lllll} & & L & & & & \\ (a_1 - 1) & 1 & 2, 4 \cdots 2^{q-2} & R & 2^{q+1} \cdots 2^{p-1} 2^p & & 2^{p+1} \cdots 2^{n-1} \\ 1 & a_2 - 1 & 2, 4 \cdots 2^{q-2} & 2^{q-1} & 2^q \cdots 2^{p-2} S & & 2^{p+1} \cdots 2^{n-1} \\ \hline a_1 & a_2 & 4, 8 \cdots 2^{q-1} & \cdots & & & 2^{p+2} \cdots 2^n \\ & & L & & & & \end{array}$$

We must now adjust the values to the right of $L - L$ to the values $a_3, a_4, \cdots, a_n$. Let us denote the sequence

$$4, 8, \cdots, 2^{q-1}, (2^{q-1} + R), 3\cdot 2^q, 3\cdot 2^{q+1}, \cdots (2^p + S), 2^{p+2}, \cdots, 2^n$$

by $\mu_1, \mu_2, \cdots, \mu_{n-2}$. Now since all the rows in the above addition are non-decreasing to the right of $L - L$, and no 1's appear, we will have proved the lemma if we can show that

$$\sum_{i=1}^{i} \mu_i \leq \sum_{i=3}^{i+3} a_i \qquad i = 1, 2, \cdots, (n-2)$$

since we have shown this to be a sufficient condition that

$$a_3, a_4, \cdots, a_n = (\mu_1, \mu_n, \cdots, \mu_{n-2})$$

and the decomposition proof we used for the first part will work. For $i \leq q - 2$, i.e., before the term $(2^{q-1} + R)$

$$\sum_{i=1}^{i} \mu_i = 4(2^i - 1)$$

and

$$\sum_{3}^{i+3} a_i \geq ia_2 \geq i2^{p-1} \geq i2^{q-1}$$

since

$$q \leq p$$

Hence

$$\sum_{1}^{i} \mu_i \leq \sum_{3}^{i+3} a_i \qquad i \leq q - 2$$

Next, for $(q - 1) \leq i \leq (p - 3)$, i.e., before the term $(2^p + S)$

$$\sum_{1}^{i} \mu_i = 4(2^{q-1} - 1) + R + 3\cdot 2^q(2^{i-q+1} - 1)$$
$$< 3\cdot 2^{i+1} - 4 \leq 3\cdot 2^{i+1} - 5$$

since

$$R < 2^q$$

also again

$$\sum_{3}^{i+3} a_i \geq i2^{p-1}$$

so that in this interval we also have the desired inequality. Finally for the last interval,

$$\sum_1^i \mu_i = 2^{i-1} - a_1 - a_2 \leq 2^{i+3} - a_1 - a_2 - 2$$

and

$$\sum_3^{i+3} a_i = \sum_1^{i+3} a_i - a_1 - a_2 \geq 2^{i+3} - a_1 - a_2 - 2$$

since

$$a_1, a_2, \cdots, a_n = (2, 4, 8, \cdots, 2^n)$$

This proves the lemma.

5. The Disjunctive Tree

It is now easy to prove the following:

Theorem 8: A disjunctive tree of n bays can be constructed with any distribution

$$a_1, a_2, \cdots, a_n = (1, 2, 4, \cdots, 2^{n-1}).$$

We may prove this by induction. We have seen it to be true for $n = 2, 3, 4$. Assuming it for n, it must be true for $n + 1$ since the Lemma shows that any

$$a_1, a_2, \cdots, a_n = (2, 4, 8, \cdots, 2^n)$$

can be decomposed into a sum which, by assumption, can be realized for the two branches of the tree.

It is clear that among the possible distributions

$$(1, 2, 4, \cdots, 2^{n-1})$$

for the tree, an "almost uniform" one can be found for all the variables but one. That is, we can distribute the load on $(n - 1)$ of them uniformly except at worst for one element. We get, in fact, for

$n = 1$	1
$n = 2$	1, 2
$n = 3$	1, 3, 3
$n = 4$	1, 4, 5, 5,
$n = 5$	1, 7, 7, 8, 8,
$n = 6$	1, 12, 12, 12, 13, 13
$n = 7$	1, 21, 21, 21, 21, 21, 21
etc.	

as nearly uniform distributions.

6. Other Distribution Problems

Now let us consider the problem of load distribution in series-parallel circuits. We shall prove the following:

Theorem 9: Any function $f(X_1, X_2, \cdots, X_n)$ may be realized with a series-parallel circuit with the following distribution:

$$(1, 2, 4, \cdots, 2^{n-2}), 2^{n-2}$$

in terms of transfer elements.

This we prove by induction. It is true for $n = 3$, since any function of three variables can be realized as follows:

$$f(X, Y, Z) = [X + f_1(Y, Z)][X' + f_2(Y, Z)]$$

and $f_1(Y, Z)$ and $f_2(Y, Z)$ can each be realized with one transfer on Y and one on Z. Thus $f(X, Y, Z)$ can be realized with the distribution 1, 2, 2. Now assuming the theorem true for $(n - 1)$ we have

$$f(X_1, X_2, \cdots, X_n) = [X_n + f_1(X_1, X_2, \cdots, X_{n-1})] [X_n' + f_2(X_1, X_2, \cdots, X_{n-1})]$$

and

$$\begin{array}{l} 2, 4, 8, \cdots, 2^{n-3} \\ 2, 4, 8, \cdots, 2^{n-3} \\ \hline 4, 8, 16, \cdots, 2^{n-2} \end{array}$$

A simple application of the Lemma thus gives the desired result. Many distributions beside those given by Theorem 9 are possible but no simple criterion has yet been found for describing them. We cannot say any distribution

$$(1, 2, 4, 8, \cdots, 2^{n-2}, 2^{n-2})$$

(at least from our analysis) since for example

$$3, 6, 6, 7 = (2, 4, 8, 8)$$

cannot be decomposed into two sets

$$a_1, a_2, a_3, a_4 = (1, 2, 4, 4)$$

and

$$b_1, b_2, b_3, b_4 = (1, 2, 4, 4)$$

It appears, however, that the almost uniform case is admissible.

As a final example in load distribution we will consider the case of a network in which a number of trees in the same variables are to be realized. A large number of such cases will be found later. The following is fairly obvious from what we have already proved.

Theorem 10: It is possible to construct m different trees in the same n variables with the following distribution:

$$a_1, a_2, \cdots, a_n = (m, 2m, 4m, \cdots, 2^{n-1}m)$$

It is interesting to note that under these conditions the bothersome 1 disappears for $m > 1$. We can equalize the load on all n of the variables, not just $n - 1$ of them, to within, at worst, one transfer element.

7. The Function $\mu(n)$

We are now in a position to study the behavior of the function $\mu(n)$. This will be done in conjunction with a treatment of the load distributions possible for the general function of n variables. We have already shown that any function of three variables can be realized with the distribution

$$1, 1, 2$$

in terms of transfer elements, and, consequently $\mu(3) \leq 4$.

Any function of four variables can be realized with the distribution

$$1, 1, (2, 4)$$

Hence $\mu(4) \leq 6$. For five variables we can get the distribution

$$1, 1, (2, 4, 8)$$

or alternatively

$$1, 5, 5, (2, 4)$$

so that $\mu(5) \leq 10$. With six variables we can get

$$1, 5, 5, (2, 4, 8) \text{ and } \mu(6) \leq 10$$

for seven,

$$1, 5, 5, (2, 4, 8, 16) \text{ and } \mu(7) \leq 16$$

etc. Also, since we can distribute uniformly on all the variables in a tree except one, it is possible to give a theorem analogous to Theorem 7 for the function $\mu(n)$:

Theorem 11: For all n

$$\mu(n) \leq \frac{2^{n+3}}{n^2}$$

For almost all n

$$\mu(n) \leq \frac{2^{n+2}}{n^2}$$

For an infinite number of n_i,

$$\mu(n) \leq (1 + \epsilon) \frac{2^{n+1}}{n^2}$$

The proof is direct and will be omitted.

PART III: SPECIAL FUNCTIONS

8. Functional Relations

We have seen that almost all functions require the order of

$$\frac{2^{n+1}}{n^2}$$

elements per relay for their realization. Yet a little experience with the circuits encountered in practice shows that this figure is much too large. In a sender, for example, where many functions are realized, some of them involving a large number of variables, the relays carry an average of perhaps 7 or 8 contacts. In fact, almost all relays encountered in practice have less than 20 elements. What is the reason for this paradox? The answer, of course, is that the functions encountered in practice are far from being a random selection. Again we have an analogue with transcendental numbers—although almost all numbers are transcendental, the chance of first encountering a transcendental number on opening a mathematics book at random is certainly much less than 1. The functions actually encountered are simpler than the general run of Boolean functions for at least two major reasons:

(1) A circuit designer has considerable freedom in the choice of functions to be realized in a given design problem, and can often choose fairly simple ones. For example, in designing translation circuits for telephone work it is common to use additive codes and also codes in which the same number of relays are operated for each possible digit. The fundamental logical simplicity of these codes reflects in a simplicity of the circuits necessary to handle them.

(2) Most of the things required of relay circuits are of a logically simple nature. The most important aspect of this simplicity is that most circuits can be broken down into a large number of small circuits. In place of realizing a function of a large number of variables, we realize many functions, each of a small number of variables, and then perhaps some function of these functions. To get an idea of the effectiveness of this consider the following example: Suppose we are to realize a function

$$f(X_1, X_2, \cdots, X_{2n})$$

of $2n$ variables. The best limit we can put on the total number of elements necessary is about $\frac{2^{2n+1}}{2n}$. However, if we know that f is a function of two functions f_1 and f_2, each involving only n of the variables, i.e. if

$$f = g(f_1, f_2)$$

$$f_1 = f_1(X_1, X_2, \cdots, X_n)$$

$$f_2 = f_2(X_{n+1}, X_{n+2}, \cdots, X_{2n})$$

then we can realize f with about

$$4 \cdot \frac{2^{n+1}}{n}$$

elements, a much lower order of infinity than $\frac{2^{2n+1}}{2n}$. If g is one of the simpler functions of two variables; for example if $g(f_1, f_2) = f_1 + f_2'$, or in any case at the cost of two additional relays, we can do still better and realize f with about $2\frac{2^{n+1}}{n}$ elements. In general, the more we can decompose a synthesis problem into a combination of simple problems, the simpler the final circuits. The significant point here is that, due to the fact that f satisfies a certain functional relation

$$f = g(f_1, f_2),$$

we can find a simple circuit for it compared to the average function of the same number of variables.

This type of functional relation may be called functional separability. It is often easily detected in the circuit requirements and can always be used to reduce the limits on the number of elements required. We will now show that most functions are not functionally separable.

Theorem 12: The fraction of all functions of n variables that can be written in the form

$$f = g(h(X_1 \cdots X_s), X_{s+1}, \cdots, X_n)$$

where $1 < s < n - 1$ approaches zero as n approaches ∞.

We can select the s variables to appear in h in $\binom{n}{s}$ ways; the function h then has 2^{2^s} possibilities and g has $2^{2^{n-s+1}}$ possibilities, since it has $n - s + 1$ arguments. The total number of functionally separable functions is therefore dominated by

$$\sum_{s=2}^{n-2} \binom{n}{s} 2^{2^s} 2^{2^{n-s+1}}$$

$$\leq (n - 3) \frac{n^2}{2} 2^{2^2} 2^{2^{n-1}}$$

and the ratio of this to $2^{2^n} \rightarrow 0$ as $n \rightarrow \infty$.

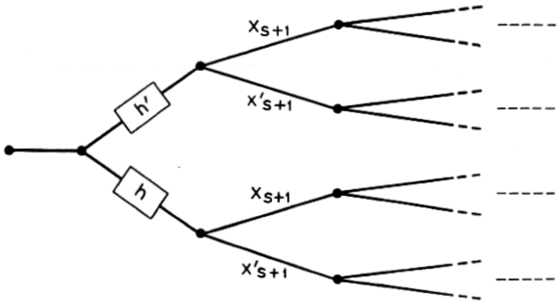


Fig. 20—Use of separability to reduce number of elements.

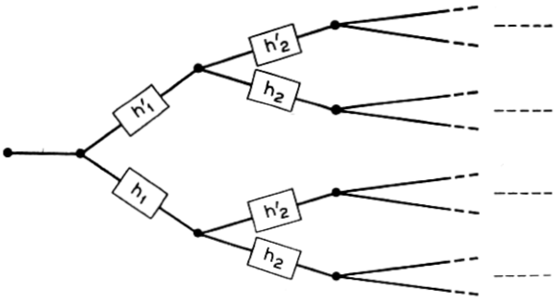


Fig. 21—Use of separability of two sets of variables.

In case such a functional separability occurs, the general design method described above can be used to advantage in many cases. This is typified by the circuit of Fig. 20. If the separability is more extensive, e.g.

$$f = g(h_1(X_1 \cdots X_s), h_2(X_{s+1} \cdots X_t), X_{t+1}, \cdots, X_n)$$

the circuit of Fig. 21 can be used, using for "h_2" either h_1 or h_2, whichever requires the least number of elements for realization together with its negative.

We will now consider a second type of functional relation which often occurs in practice and aids in economical realization. This type of relation may be called group invariance and a special case of it, functions symmetric in all variables, has been considered in (6). A function $f(X_1, \cdots, X_n)$ will be said to be symmetric in X_1, X_2 if it satisfies the relation

$$f(X_1, X_2, \cdots, X_n) = f(X_2, X_1, \cdots, X_n).$$

It is symmetric in X_1 and X'_2 if it satisfies the equation

$$f(X_1, X_2, \cdots, X_n) = f(X'_2, X'_1, X_3, \cdots, X_n)$$

These also are special cases of the type of functional relationships we will consider. Let us denote by

$N_{oo \cdots o} = I$ the operation of leaving the variables in a function as they are,

$N_{1oo \cdots o}$ the operation of negating the first variable (i.e. the one occupying the first position),

$N_{o1o \cdots o}$ that of negating the second variable,

$N_{11o \cdots o}$ that of negating the first two, etc.

So that $N_{101}f(X, Y, Z) = f(X'YZ')$ etc.

The symbols Ni form an abelian group, with the important property that each element is its own inverse; $NiNi = I$ The product of two elements may be easily found — if $N_i N_j = N_k$, k is the number found by adding i and j as though they were numbers in the base two but *without carrying*.

Note that there are 2^n elements to this "negating" group. Now let

$S_{1,2,3,\ldots,n} = I =$ the operation of leaving the variables of a function in the same order

$S_{2,1,3,\ldots,n} =$ be that of interchanging the first two variables

$S_{3,2,1,4,\ldots,n} =$ that of inverting the order of the first three, etc.

Thus

$$S_{312}f(X, Y, Z) = f(Z, X, Y)$$

$$S_{312}f(Z, X, Y) = S_{312}^2 f(X, Y, Z) = f(Y, Z, X)$$

etc. The S_i also form a group, the famous "substitution" or "symmetric" group. It is of order $n!$. It does not, however, have the simple properties of the negating group—it is not abelean ($n > 2$) nor does it have the self inverse property.* The negating group is not cyclic if $n > 2$, the symmetric group is not if $n > 3$.

The outer product of these two groups forms a group G whose general element is of the form $N_i S_j$ and since i may assume 2^n values and j, $n!$ values, the order of G is $2^n n!$

It is easily seen that $S_j N_i = N_k S_j$, where k may be obtained by per-

* This is redundant; the self inverse property implies commutativity for if $XX = I$ then $XY = (XY)^{-1} = Y^{-1}X^{-1} = YX$.

forming on i, considered as an ordered sequence of zero's and one's, the permutation S_j. Thus

$$S_{2314}N_{1100} = N_{1010}S_{2314}.$$

By this rule any product such as $N_i S_j N_k N_l S_m N_n S_p$ can be reduced to the form

$$N_i N_j \cdots N_n S_p S_q \cdots S_r$$

and this can then be reduced to the standard form $N_i S_j$.

A function f will be said to have a non-trivial group invariance if there are elements $N_i S_j$ of G other than I such that identically in all variables

$$N_i S_j f = f.$$

It is evident that the set of all such elements, $N_i S_j$, for a given function, forms a subgroup G_1 of G, since the product of two such elements is an element, the inverse of such an element is an element, and all functions are invariant under I.

A group operator leaving a function f invariant implies certain equalities among the terms appearing in the expanded form of f. To show this, consider a fixed $N_i S_j$, which changes in some way the variables (say) $X_1, X_2, \cdots, X_r$. Let the function $f(X_1, \cdots, X_n)$ be expanded about $X_1, \cdots, X_r$:

$$\begin{aligned} f = {} & [X_1 + X_2 + \cdots + X_r + f_1(X_{r+1}, \cdots, X_n)] \\ & [X_1' + X_2 + \cdots + X_r + f_2(X_{r+1}, \cdots, X_n)] \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & [X_1' + X_2' + \cdots + X_r' + f_{2^r}(X_{r+1}, \cdots, X_n)] \end{aligned}$$

If f satisfies $N_i S_j f = f$ we will show that there are at least $\frac{1}{4}2^r$ equalities between the functions $f_1, f_2, \cdots, f_{2^r}$. Thus the number of functions satisfying this relation is

$$\leq (2^{2^{n-r}})^{\frac{3}{4}2^r} = 2^{\frac{3}{4}2^n}$$

since each independent f_i can be any of just $2^{2^{n-r}}$ functions, and there are at most $\frac{3}{4}\, 2^r$ independent ones. Suppose $N_i S_j$ changes

$$X_1, X_2, \cdots, X_r \qquad \text{A}$$

into

$$X_{a_1}^*, X_{a_2}^*, \cdots, X_{a_r}^* \qquad \text{B}$$

where the *'s may be either primes or non primes, but no $X_{a_i}^* = X_i$. Give

X_1 the value 0. This fixes some element in B namely, X_{a_i} where $a_i = 1$. There are two cases:
(1) If this element is the first term, $a_1 = 1$, then we have

$$0 X_2, \cdots, X_r$$

$$1 X_{a_1}, \cdots, X_{a_r}$$

Letting $X_2, \cdots, X_r$ range through their 2^{r-1} possible sets of values gives 2^{r-1} equalities between different functions of the set f_i since these are really

$$f(X_1, X_2, \cdots, X_r, X_{r+1}, \cdots, X_n)$$

with $X_1, X_2, \cdots, X_r$ fixed at a definite set of values.
(2) If the element in question is another term, say X_{a_2}, we then give X_2 in line A the opposite value, $X_2 = (X_{a_2}^*)' = (X_2^*)'$. Now proceeding as before with the remaining $r - 2$ variables we establish 2^{r-2} equalities between the f_i.

Now there are not more* than $2^n n!$ relations

$$N_i S_j f = f$$

of the group invariant type that a function could satisfy, so that the number of functions satisfying *any* non-trivial relation

$$\leq 2^n\, n!\, 2^{\frac{3}{4}2^n}.$$

Since

$$2^n\, n!\, 2^{\frac{3}{4}2^n}/2^{2^n} \rightarrow 0 \qquad \text{as } n \rightarrow \infty$$

we have:

Theorem 13: Almost all functions have no non-trivial group invariance.

It appears from Theorems 12 and 13 and from other results that almost all functions are of an extremely chaotic nature, exhibiting no symmetries or functional relations of any kind. This result might be anticipated from the fact that such relations generally lead to a considerable reduction in the number of elements required, and we have seen that almost all functions are fairly high in "complexity".

If we are synthesizing a function by the disjunctive tree method and the function has a group invariance involving the variables

$$X_1, X_2, \cdots, X_r$$

at least 2^{r-2} of the terminals in the corresponding tree can be connected to

* Our factor is really less than this because, first, we must exclude $N_i S_j = I$; and second, except for self inverse elements, one relation of this type implies others, viz. the powers $(N_i S_j)^p f = f$.

other ones, since at least this many equalities exist between the functions to be joined to these terminals. This will, in general, produce a considerable reduction in the contact requirements on the remaining variables. Also an economy can usually be achieved in the M network. In order to apply this

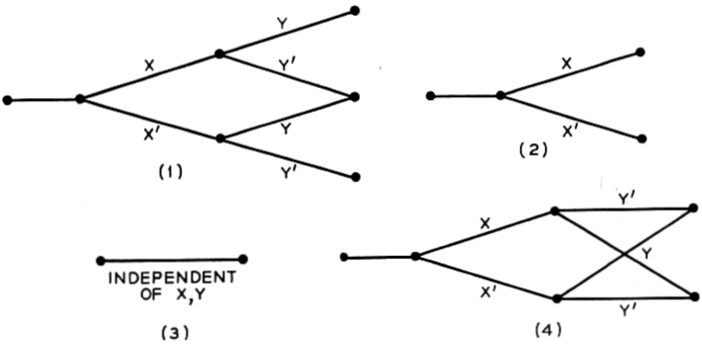


Fig. 22—Networks for group invariance in two variables.

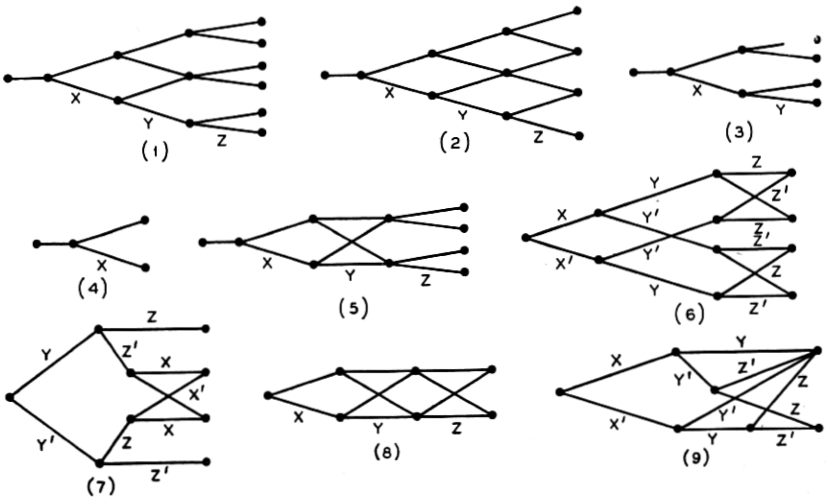


Fig. 23—Networks for group invariance in three variables.

method of design, however, it is essential that we have a method of determining which, if any, of the $N_i S_j$ leave a function unchanged. The following theorem, although not all that might be hoped for, shows that we don't need to evaluate $N_i S_j f$ for all $N_i S_j$ but only the $N_i f$ and $S_j f$.

Theorem 14: A necessary and sufficient condition that $N_i S_j f = f$ *is that* $N_i f = S_j f$.

This follows immediately from the self inverse property of the N_i. Of course, group invariance can often be recognized directly from circuit requirements in a design problem.

Tables I and II have been constructed for cases where a relation exists involving two or three variables. To illustrate their use, suppose we have a function such that

$$N_{111} S_{312} f = f$$

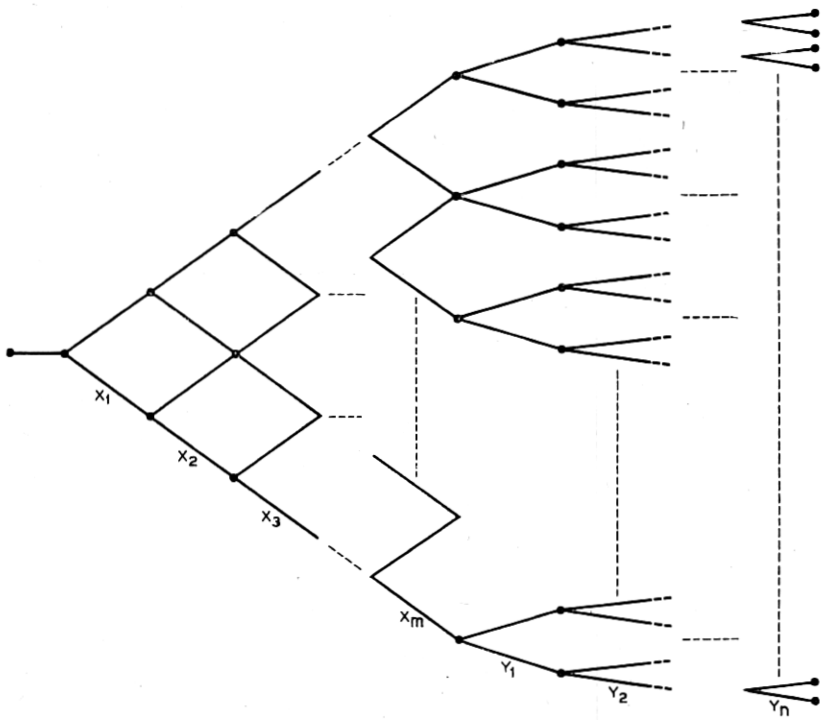


Fig. 24—*M* network for partially symmetric functions.

The corresponding entry $Z'Y'X$ in the group table refers us to circuit 9 of Fig. 23. The asterisk shows that the circuit may be used directly; if there is no asterisk an interchange of variables is required. We expand f about X, Y, Z and only two different functions will appear in the factors. These two functions are realized with two trees extending from the terminals of the network 9. Any such function f can be realized with (using just one variable in the N network)

$$9 + 2(2^{n-4} - 2) + 2$$

$$= 2^{n-3} + 7 \qquad \text{elements,}$$

a much better limit than the corresponding

$$2^{n-1} + 18$$

for the general function.

TABLE I

GROUP INVARIANCE INVOLVING TWO VARIABLES (SUPERSCRIPTS REFER TO FIG. 22)

	S_{12}	S_{21}
N_{oo}	$(x\ y)$	$(y\ x)^{1*}$
N_{o1}	$(x\ y')^{2*}$	$(y\ x')^{3*}$
N_{10}	$(x'y)^{2}$	$(y'x)^{3*}$
N_{11}	$(x'y')^{4}$	$(y'x')^{1}$

TABLE II

GROUP INVARIANCE INVOLVING THREE VARIABLES (SUPERSCRIPTS REFER TO FIG. 23)

	S_{123}	S_{132}	S_{213}	S_{231}	S_{312}	S_{321}
N_{ooo}	XYZ	XZY^{1}	YXZ^{1*}	YZX^{2*}	ZXY^{2*}	ZYX [illegible]
N_{oo1}	XYZ'^{3*}	XZY'^{4*}	YXZ'^{7}	YZX'^{9}	ZXY'^{9}	ZYX'^{4}
N_{o1o}	$XY'Z^{3}$	$XZ'Y^{4*}$	$YX'Z^{4}$	$YZ'X^{9}$	$ZX'Y^{9}$	$ZY'X^{7}$
N_{o11}	$XY'Z'^{5}$	$XZ'Y'^{1}$	$YX'Z'^{8}$	$YZ'X'^{2}$	$ZX'Y'^{2}$	$ZY'X'^{8}$
N_{1oo}	$X'YZ^{3}$	$X'ZY^{7*}$	$Y'XZ^{4}$	$Y'ZX^{9}$	$Z'XY^{9}$	$Z'YX^{4}$
N_{101}	$X'YZ'^{5}$	$X'ZY'^{8*}$	$Y'XZ'^{8*}$	$Y'ZX'^{2}$	$Z'XY'^{2}$	$Z'YX'^{1}$
N_{110}	$X'Y'Z^{5*}$	$X'Z'Y^{8*}$	$Y'X'Z^{1}$	$Y'Z'X^{2}$	$Z'X'Y^{2}$	$Z'Y'X^{8*}$
N_{111}	$X'Y'Z'^{6}$	$X'Z'Y'^{7}$	$Y'X'Z'^{7}$	$Y'Z'X'^{9}$	$Z'X'Y'^{9}$	$Z'Y'X'^{7}$

9. PARTIALLY SYMMETRIC FUNCTIONS

We will say that a function is "partially symmetric" or "symmetric in a certain set of variables" if these variables may be interchanged at will without altering the function. Thus

$$XYZW + (XY' + X'Y)W + WZ'$$

is symmetric in X and Y. Partial symmetry is evidently a special case of the general group invariance we have been considering. It is known that any function symmetric in all variables can be realized with not more than n^2 elements, where n is the number of variables.[6] In this section we will improve and generalize this result.

Theorem 15: Any function $f(X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots Y_n)$ *symmetric in* $X_1, X_2, \cdots, X_m$ *can be written*

$$\begin{aligned} &f(X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_n) \\ &= [S_o(X_1, X_2, \cdots, X_m) + f_o(Y_1, Y_2, \cdots, Y_n)]. \\ &\quad [S_1(X_1, X_2, \cdots, X_m) + f_1(Y_1, Y_2, \cdots, Y_n)]. \\ &\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\quad [S_m(X_1, X_2, \cdots, X_m) + f_m(Y_1, Y_2, \cdots, Y_n)] \qquad (6) \end{aligned}$$

where

$$\begin{aligned} &f_K(Y_1, Y_2, \cdots, Y_n) \\ &= f(\underbrace{0, 0, \cdots, 0}_{k\ 0\text{'s}}, \underbrace{1, 1, \cdots, 1}_{(m-k)\ 1\text{'s}}, Y_1, Y_2, \cdots, Y_n) \end{aligned}$$

and $S_k(X_1, X_2, \cdots, X_m)$ *is the symmetric function of* $X_1, X_2, \cdots, X_n$ *with k for its only a-number.*

This theorem follows from the fact that since f is symmetric in $X_1, X_2, \cdots, X_m$ the value of f depends only on the *number* of X's that are zero and the values of the Y's. If exactly K of the X's are zero the value of f is therefore f_K, but the right-hand side of (6) reduces to f_K in this case, since then $S_j(X_1, X_2, \cdots, X_m) = 1, j \neq K$, and $S_K = 0$.

The expansion (6) is of a form suitable for our design method. We can realize the disjunctive functions $S_K(X_1, X_2, \cdots, X_n)$ with the symmetric function lattice and continue with the general tree network as in Fig. 24, one tree from each level of the symmetric function network. Stopping the trees at Y_{n-1}, it is clear that the entire network is disjunctive and a second application of Theorem 1 allows us to complete the function f with two elements from Y_n. Thus we have

Theorem 16. Any function of $m + n$ *variables symmetric in* m *of them can be realized with not more than the smaller of*

$$(m + 1)(\lambda(n) + m) \text{ or } (m + 1)(2^n + m - 2) + 2$$

elements. In particular a function of n *variables symmetric in* $n - 2$ *or more of them can be realized with not more than*

$$n^2 - n + 2$$

elements.

If the function is symmetric in $X_1, X_2, \cdots, X_m$, and also in $Y_1, Y_2, \cdots, Y_r$, and not in $Z_1, Z_2, \cdots, Z_n$ it may be realized by the same method, using symmetric function networks in place of trees for the Y variables. It should be expanded first about the X's (assuming $m < r$) then about the Y's and finally the Z's. The Z part will be a set of $(m + 1)(r + 1)$ trees.

References

1. G. Birkhoff and S. MacLane, "A Survey of Modern Algebra," Macmillan, 1941.
2. L. Couturat, "The Algebra of Logic," Open Court, 1914.
3. J. H. Woodger, "The Axiomatic Method in Biology," Cambridge, 1937.
4. W. S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bull. Math. Biophysics*, V. 5, p. 115, 1943.
5. E. C. Berkeley, "Boolean Algebra and Applications to Insurance," *Record (American Institute of Actuaries)*, V. 26, p. 373, 1947.
6. C. E. Shannon, "A Symbolic Analysis of Relay and Switching Circuits," *Trans. A. I. E. E.*, V. 57, p. 713, 1938.
7. J. Riordan and C. E. Shannon, "The Number of Two-Terminal Series Parallel Networks," *Journal of Mathematics and Physics*, V. 21, No. 2, p. 83, 1942.
8. A. Nakashima, Various papers in *Nippon Electrical Communication Engineering*, April, Sept., Nov., Dec., 1938.
9. H. Piesch, Papers in Archiv. from *Electrotechnik* XXXIII, p. 692 and p. 733, 1939.
10. G. A. Montgomerie, "Sketch for an Algebra of Relay and Contactor Circuits," *Jour. I. of E. E.*, V. 95, Part III, No. 36, July 1948, p. 303.
11. G. Pólya, "Sur Les Types des Propositions Composées," *Journal of Symbolic Logic*, V. 5, No. 3, p. 98, 1940.